



Published in final edited form as:

Commun Stat Theory Methods. 2017 ; 46(1): 344–353. doi:10.1080/03610926.2014.991040.

Sample Size Calculations for Time-Averaged Difference of Longitudinal Binary Outcomes

Ying Lou,

Department of Statistical Science, Southern Methodist University, Dallas, TX

Jing Cao,

Department of Statistical Science, Southern Methodist University, Dallas, TX

Song Zhang, and

Department of Clinical Sciences, UT Southwestern Medical Center, Dallas, TX

Chul Ahn

Department of Clinical Sciences, UT Southwestern Medical Center, Dallas, TX

Abstract

In clinical trials with repeated measurements, the responses from each subject are measured multiple times during the study period. Two approaches have been widely used to assess the treatment effect, one that compares the rate of change between two groups and the other that tests the time-averaged difference (TAD). While sample size calculations based on comparing the rate of change between two groups have been reported by many investigators, the literature has paid relatively little attention to the sample size estimation for time-averaged difference (TAD) in the presence of heterogeneous correlation structure and missing data in repeated measurement studies. In this study we investigate sample size calculation for the comparison of time-averaged responses between treatment groups in clinical trials with longitudinally observed binary outcomes. The GEE approach is used to derive a closed-form sample size formula, which is flexible enough to account for arbitrary missing patterns and correlation structures. In particular, we demonstrate that the proposed sample size can accommodate a mixture of missing patterns, which is frequently encountered by practitioners in clinical trials. To our knowledge, this is the first study that considers the mixture of missing patterns in sample size calculation. Our simulation shows that the nominal power and type I error are well preserved over a wide range of design parameters. Sample size calculation is illustrated through an example.

Keywords

sample size; GEE; binary; time-averaged differences; repeated measurements; mixture of missing patterns

1 Introduction

Diggle et al. (2002) provided sample size formulas to assess the treatment effect in repeated measurement studies assuming no missing data and the compound symmetry (CS) correlation structure among outcomes from the same subject. One sample size formula is based on comparing the rate of change between two groups and the other is based on comparing the time-averaged responses between two groups. The rate-of-change approach has been widely used to investigate if the rates of changes in outcomes are significantly different over the study period between two treatment groups. The time-averaged difference (TAD) is particularly meaningful in cases where the treatment effect has rapid onset and repeated measurements continue across an extended period after a maximum effect is achieved. In these situations, the TAD across time between two treatment groups may provide a more powerful evaluation of treatment efficacy than the within-subject trends or change scores (Overall & Doyle 1994). The TAD approach is also often used when outcome varies with time. The precision of experiment is increased by taking multiple measurements from each study subject. The TAD approach is used by incorporating the correlation among measurements from the same subject to determine if there is an overall difference in the proportion of responders between two treatment groups over the study period. In this study, we will investigate the sample size requirement for clinical trials with repeatedly measured binary outcomes, where the treatment effect is assessed based on TAD between two groups.

By obtaining multiple outcome measurements from the same subjects, researchers hope to reduce intra-patient variability and thus increase study power. The challenge in designing a clinical trial with repeated measurements, however, is that the impact of within-subject correlation and missing data needs to be taken into account appropriately. Lui (1991) derived a sample size formula for repeated binary outcomes with the Markov dependency given subject-specific probabilities without incorporating missing data. Lipsitz & Fitzmaurice (1994) used the weighted least squares approach to estimate sample size for the detection of a clinically important treatment effect in repeated measurement studies with a binary response. The generalized estimating equation (GEE) method (Zeger & Liang 1986) has been widely used to make inference based on repeated measurements. Under the MCAR (missing completely at random) assumption, it provides consistent estimators for regression parameters and their variance-covariance matrix even when the correlation structure is misspecified. Liu & Liang (1997) presented a general method to estimate sample size for correlated observations using GEE. Treating repeated binary outcomes as a special case, they provided a closed-form sample size formula for the test of time-averaged difference, but their method did not account for missing data either. Patel & Rowe (1999) presented sample size formulas for binary and count outcomes using GEE without consideration of missing data. Jung & Ahn (2005) provided a closed-form sample size formula for comparing the rates of change in repeated binary measurements with a logit link function. This formula is flexible to account for missing data and various correlation structures.

In this paper, we investigate sample size calculation for detecting TAD between two groups based on longitudinally measured binary outcomes. Care must be taken in derivation because of the correlation introduced when several measurements are taken from the same individual. The correlation structures may take on several forms depending on the nature of

the experiment and the subjects involved. This procedure allows us to calculate sample sizes and power under arbitrary correlation structures, including compound symmetry and AR(1), etc.

In Section 2, we briefly review the GEE method for binary repeated measurements. In Section 3 we present a closed-form sample size formula for the evaluation of TAD that is flexible enough to accommodate arbitrary missing patterns and correlation structures. We evaluate the sample size formula through numerical simulation under various settings in Section 4. Finally, we apply our sample size formula to a real data example for illustration in Section 5. The final section is devoted to discussion.

2 Generalized Estimating Equation Estimator

Let Y_{ij} be the binary response obtained at time t_{ij} ($j = 1, \dots, m$) from subject i ($i = 1, \dots, n$). We use $r_i = 1/0$ to indicate that subject i belongs to the treatment/control group, and $\bar{r} = E(r_i)$ is the proportion of subjects randomly assigned to the treatment group. To evaluate TAD between two groups, we model Y_{ij} with the following logistic model: $Y_{ij} \sim \text{Bernoulli}(p_{ij})$ and

$$\log \left(\frac{p_{ij}}{1 - p_{ij}} \right) = \beta_1 + \beta_2 r_i, \quad \text{for } j=1, \dots, m. \quad (1)$$

Here β_1 models the time-averaged response on the log-odds scale for the control group, and β_2 is the log odds ratio between treatment and control, representing the treatment effect. Our primary interest is to test the null hypothesis $H_0 : \beta_2 = 0$. To facilitate later derivation, we reparameterize Equation (1) as

$$\log \left(\frac{p_{ij}}{1 - p_{ij}} \right) = b_1 + b_2 (r_i - \bar{r}), \quad (2)$$

where $b_1 \equiv \beta_1 + \beta_2 \bar{r}$, and $b_2 \equiv \beta_2$. Hence, testing $b_2 = 0$ is equivalent to testing $\beta_2 = 0$. From Equation (2) we have

$$p_{ij}(\mathbf{b}) = \frac{e^{\mathbf{b}' Z_{ij}}}{1 + e^{\mathbf{b}' Z_{ij}}}, \quad (3)$$

where $\mathbf{b} = (b_1, b_2)'$ and $Z_{ij} = (1, r_i - \bar{r})'$.

According to Liang & Zeger (1986), a GEE estimator $\hat{\mathbf{b}}$ is obtained by solving $U_n(\mathbf{b}) = 0$, where

$$U_n(\mathbf{b}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{j=1}^m [y_{ij} - p_{ij}(\mathbf{b})] Z_{ij}. \quad (4)$$

The equation is solved using the Newton-Raphson algorithm: at the l th iteration,

$$\hat{\mathbf{b}}^{(l)} = \hat{\mathbf{b}}^{(l-1)} + n^{-1/2} \mathbf{A}_n^{-1} \left(\hat{\mathbf{b}}^{(l-1)} \right) U_n \left(\hat{\mathbf{b}}^{(l-1)} \right), \quad (5)$$

where

$$\mathbf{A}_n(\mathbf{b}) = -n^{-1/2} \frac{\partial U_n(\mathbf{b})}{\partial \mathbf{b}} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m p_{ij} q_{ij} \begin{pmatrix} 1 & r_i - \bar{r} \\ r_i - \bar{r} & (r_i - \bar{r})^2 \end{pmatrix}. \quad (6)$$

Here $q_{ij} = 1 - p_{ij}$. Liang & Zeger (1986) showed that $n(\hat{\mathbf{b}} - \mathbf{b})$ approximately follows an $N(\mathbf{0}, \mathbf{V})$ distribution as $n \rightarrow \infty$, and \mathbf{V} can be consistently estimated by

$\mathbf{V}_n = \mathbf{A}_n^{-1}(\hat{\mathbf{b}}) \Sigma_n(\hat{\mathbf{b}}) \mathbf{A}_n^{-1}(\hat{\mathbf{b}})$, with

$$\Sigma_n(\hat{\mathbf{b}}) = \frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^m \hat{\epsilon}_{ij} Z_{ij} \right)^{\otimes 2}. \quad (7)$$

Here $\hat{\epsilon}_{ij} = y_{ij} - p_{ij}(\hat{\mathbf{b}})$, and $\mathbf{c}^{\otimes 2} = \mathbf{c}\mathbf{c}^T$ for a vector \mathbf{c} . To make inference about TAD between two groups, we reject the null hypothesis $H_0: b_2 = 0$ if $|\sqrt{n} \hat{b}_2 / \hat{\sigma}_2| > z_{1-\alpha/2}$, where $\hat{\sigma}_2^2$ is the (2, 2)th element of \mathbf{V}_n and $z_{1-\alpha/2}$ is the 100(1 - α /2)th percentile of the standard normal distribution.

3 A Closed Form Sample size Formula

At the experimental design stage, researchers need to determine how many subjects are needed such that the trial can detect a treatment effect (TAD) of β_{20} with a power $1 - \gamma$ at a significance level of α . Let \mathbf{A} and Σ denote the limits of \mathbf{A}_n and Σ_n , respectively. As $n \rightarrow \infty$, \mathbf{V}_n converges to $\mathbf{V} = \mathbf{A}^{-1} \Sigma \mathbf{A}^{-1}$. Let σ_2^2 be the (2, 2)th element of \mathbf{V} . The required sample size is

$$n = \frac{\sigma_2^2 (z_{1-\alpha/2} + z_{1-\gamma})^2}{\beta_{20}^2}. \quad (8)$$

In the above derivation we have assumed no missing data. i.e., all subjects have complete measurements at (t_1, \dots, t_m) . Missing data in clinical trials is a challenge frequently encountered by researchers, where subjects fail to provide measurements at scheduled time points for various reasons, such as equipment malfunction, schedule conflict, input error, and dropout from study, etc. In the following, we show that a closed-form sample size formula can be derived under the MCAR (missing completely at random) assumption. To accommodate the possible presence of missing data, we use $\delta_{ij} = 1/0$ to indicate whether outcome Y_{ij} is observed/missing. First we have general formulae of $A_n(\hat{\mathbf{b}})$ and $\Sigma_n(\hat{\mathbf{b}})$ that accommodate missing data:

$$A_n(\hat{\mathbf{b}}) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \Delta_{ij} p_{ij}(\hat{\mathbf{b}}) q_{ij}(\hat{\mathbf{b}}) \begin{pmatrix} 1 & r_i - \bar{r} \\ r_i - \bar{r} & (r_i - \bar{r})^2 \end{pmatrix},$$

$$\Sigma_n(\hat{\mathbf{b}}) = \frac{1}{n} \sum_{i=1}^n \left(\sum_{j=1}^m \Delta_{ij} \hat{\epsilon}_{ij} Z_{ij} \right)^{\otimes 2}.$$

To facilitate discussion, we introduce a few more notations. We define $p_1 = e^{\beta_1}/(1+e^{\beta_1})$ and $p_2 = e^{\beta_1+\beta_2}/(1+e^{\beta_1+\beta_2})$ to be the true response rates in the control and treatment groups, respectively. Similarly we define $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$. We use $\rho_{jj'} = \text{corr}(Y_{ij}, Y_{ij'})$ to denote within-subject correlation, with $\rho_{jj} = 1$. Finally, we define $\delta_j = E(\delta_{ij})$ to be the proportion of subjects with observations at t_j , and $\delta_{jj'} = E(\delta_{ij} \delta_{ij'})$ to be the proportion of subjects with observations at t_j and $t_{j'}$ simultaneously ($j \neq j'$). Note that $\delta_{jj} = \delta_j$.

Theorem 1

In the presence of missing data, as $n \rightarrow \infty$, $V_n = A_n^{-1}(\hat{\mathbf{b}}) \Sigma_n(\hat{\mathbf{b}}) A_n^{-1}(\hat{\mathbf{b}}) \rightarrow V$, and the (2,2)th element of V has a closed form

$$\sigma_2^2 = \frac{\tau \sum_{j=1}^m \sum_{j'=1}^m \delta_{jj'} \rho_{jj'}}{\left(\sum_{j=1}^m \delta_j\right)^2 \sigma_r^2 p_1 q_1 p_2 q_2} \tag{9}$$

where $\tau = (1 - \bar{r})p_1q_1 + \bar{r}p_2q_2$ and $\sigma_r^2 = \bar{r} \left(1 - \bar{r}\right)$.

Proof—For proof, see Appendix A.

Here τ is effectively the pooled variance from the control and treatment groups. The general sample size formula which accounts for missing data can be obtained by plugging Equation (9) into Equation (8). Note that in Equation (9) missing data is taken into account through the specification of δ_j and $\delta_{jj'}$, and within-subject correlation is taken into account through the specification of $\rho_{jj'}$. It is flexible enough to accommodate arbitrary types of missing pattern and correlation. The other factors that affect sample size requirement include the true TAD between two groups represented by β_{20} , the randomization ratio represented by \bar{r} , and the baseline response rate represented by β_{10} . For a continuous outcome, sample size

requirement is usually independent of the control mean. For a binary outcome, however, variance and mean (response rate) are associated, $\text{Var}(Y) = p(1 - p)$, thus the baseline response rate should be included as a designing factor in sample size calculation.

In clinical trials, researchers usually encounter two types of missing patterns: One called independent missing (IM), where a patient can have missing values independently over (t_1, \dots, t_m) , thus $\delta_{jj'} = \delta_j \delta_{j'}$ for $j < j'$. Note that $\delta_{jj} = \delta_j$. The other called monotone missing (MM), where a patient missing the measurement at t_j will miss all subsequent measurements, thus $\delta_{jj'} = \delta_{j'}$ for $j < j'$. Intuitively, we might consider MM a more serious type of missing problem because missing values tend to be concentrated in a group of subjects who drop out of the study permanently. From Equation (9) we can mathematically confirm this intuition because given the same set of marginal observing probabilities $(\delta_1, \dots, \delta_m)$, the joint probability under IM ($\delta_{jj'} = \delta_j \delta_{j'}$) is always smaller than that under MM ($\delta_{jj'} = \delta_{j'}$, $j < j'$). The larger variance (σ_2^2) under MM leads to a larger sample size requirement.

It is also likely that missing values occur following a mixed pattern, denoted as MIX. For example, some patients might drop out of study while others might miss a few appointments randomly over the study period. Equation (9) is flexible enough to accommodate this mixture type of missing patterns. Let $(\delta_1^{(IM)}, \dots, \delta_m^{(IM)})$ and $(\delta_1^{(MM)}, \dots, \delta_m^{(MM)})$ be the marginal observant probabilities under the IM and MM patterns, respectively. It is likely that patients under different missing patterns have different marginal probabilities. We also use $\delta_{jj'}^{(IM)}$ and $\delta_{jj'}^{(MM)}$ to denote the corresponding joint probabilities under each pattern, as described above. Suppose in a clinical trial the proportions of patients who would potentially follow the IM and MM patterns are w and $(1 - w)$, respectively. Then a more general sample size formula to accommodate a mixture of missing patterns can be obtained by replacing δ_j and $\delta_{jj'}$ in Equation (9) with $\delta_j^{(MIX)} = w\delta_j^{(IM)} + (1 - w)\delta_j^{(MM)}$ and $\delta_{jj'}^{(MIX)} = w\delta_{jj'}^{(IM)} + (1 - w)\delta_{jj'}^{(MM)}$. Here the superscript (MIX) indicates that the marginal and joint probabilities are calculated under the mixture of missing patterns.

4 Simulation Studies

We conduct simulations to evaluate the performance of the proposed sample size formula under various design configurations. We set $m = 6$. Without loss of generality, we assume that responses are assessed at equidistant time points, where $t_j = (j - 1)$, $j = 1, \dots, 6$. We consider two within-subject correlation structures: One is compound symmetric (CS) which assumes equal correlation regardless of temporal distance between measurements, $\rho_{jj'} = \rho$ ($j < j'$); The other is the AR(1) structure, where correlation decreases as the distance between measurements increases, $\rho_{jj'} = \rho^{|t_j - t_{j'}|}$. It is obvious from Equation (9) that sample size requirement increases with the value of ρ under either CS or AR(1). We consider $\rho = 0.3$ and 0.5 in simulation. We also explore three missing patterns: IM, MM, and MIX. We assume equal marginal probabilities for IM and MM, denoted by $\boldsymbol{\delta} = (\delta_1, \dots, \delta_m)'$, with four scenarios:

$$\begin{aligned} \delta_1 &= (1, 1, 1, 1, 1, 1) , \\ \delta_2 &= (1, 0.95, 0.9, 0.85, 0.8, 0.75) , \\ \delta_3 &= (1, 0.99, 0.96, 0.91, 0.84, 0.75) , \\ \delta_4 &= (1, 0.91, 0.84, 0.79, 0.76, 0.75) . \end{aligned}$$

Under δ_1 all patients provide complete observations. Under δ_2 the probability of patients providing a measurement decrease by 5% at each subsequent time point. Under δ_3 the observing probability decreases mildly at the beginning of the experiment and decreases sharply at the end of the experiment. The fourth scenario δ_4 present a missing pattern which is opposite of scenario δ_3 . It has a bigger portion of observation missed at the beginning of the experiment than that at the end of the experiment. We further set $w = 0.5$ for the MIX pattern. The nominal levels of type I error and power are set at $\alpha = 0.05$, and $1 - \gamma = 0.8$, respectively. We set $r_1 = 0.5$ which implies a balanced design. We also consider two levels of control response rate $p_1 = 0.5$ and 0.2 (equivalently, $\beta_1 = 0$ and -1.39). We set the true treatment effect at $\beta_2 = 0.5$.

For every combination of the aforementioned design factors (correlation structure, correlation parameter ρ , missing pattern, marginal probability δ , control response rate p_1), the simulation study is carried out as follows:

1. Calculate sample size (n) based on the proposed sample size formula;
2. For iteration $l = 1, \dots, L$ ($L = 5000$),
 - (a) Simulate a null data set (under $\beta_2 = 0$) and an alternative data set (under $\beta_2 = 0.5$), each with n subjects. Every subject has a binary vector of measurements, $\mathbf{Y}_j = (Y_{j1}, \dots, Y_{jm})'$, with mean determined by (β_1, β_2) and within-subject correlation $\rho_{jj'}$ determined by ρ and the corresponding correlation structure. Generation of correlated binary vectors is based on the algorithm of Emrich & Piedmonte (1991).
 - (b) Generate missing values according to marginal probability δ and the specified missing pattern.
 - (c) Calculate $\hat{\beta}_2, \hat{\sigma}_2^2$ using Equations (5),(6), and (7), denoted as $\hat{\beta}_2^{(l,0)}$ and $\hat{\sigma}_2^{2(l,0)}$ for the null data set and $\hat{\beta}_2^{(l,1)}$ and $\hat{\sigma}_2^{2(l,1)}$ for the alternative data set. Here the superscripts ($l, 0/1$) indicate that the estimators are obtained based on the l th null/alternative data set.
3. Estimate the empirical type I error by $\sum_{l=1}^L I \left(|\hat{\beta}_2^{(l,0)}| / \sqrt{\hat{\sigma}_2^{2(l,0)}} > z_{1-\alpha/2} \right) / L$ and the empirical power by $\sum_{l=1}^L I \left(|\hat{\beta}_2^{(l,1)}| / \sqrt{\hat{\sigma}_2^{2(l,1)}} > z_{1-\alpha/2} \right) / L$.

In Tables 1 and 2 we present the sample sizes (together with the empirical powers, empirical type I errors) calculated under $p_1 = 0.5$ and $p_1 = 0.2$, respectively. Note that the results under complete data (δ_1) are redundant for the MM and MIX missing patterns and thus omitted. We have several observations: 1) Other design parameters being the same, sample size

requirement increases with the correlation parameter ρ , which is obvious from Equation (9); 2) The MM pattern causes a greater information loss than IM, which leads to a larger sample size requirement. Furthermore, the sample size under MIX is always between those under IM and MM; 3) Although $\delta_2 - \delta_4$ have equal dropout rate at the end of study, their sample size requirements are different due to the different distributions of missing values during follow up; 4) Tables 1 and 2 show that the empirical powers and type I errors are preserved close to their nominal levels over a wide range of sample sizes (143 to 460). 5) Tables 1 and 2 demonstrate that the level of control response rate (p_1) can have a great impact on sample size requirement. However, from Equation (9) we can see that the relationship between p_1 and σ_2^2 (or sample size n) is not straightforward.

In real clinical trials, the outcome measurements are rarely obtained at the exact scheduled time. To evaluate the robustness of the proposed sample size method, we consider a more realistic scenario where responses are measured at random time points. Specifically, the sample size is calculated assuming that $t_1 = 0$, $t_6 = 5$, and t_j is uniformly distributed in $(t_j - 1/2, t_j + 1/2)$ for $j = 2, \dots, 5$. The calculated sample sizes are the same as those in Tables 1 and 2. We observe that the empirical powers and type I errors are generally close to their nominal levels (Tables not shown here), suggesting that the proposed sample size performs well even when the fixed-time assumption is violated in reality.

5 Example

We apply the proposed sample size method to the example in PASS sample size software manual (Hintze 2013). In order to determine the efficacy of a prophylactic treatment for the common cold, subjects will be randomly assigned to a treatment group or a placebo group with an equal probability, and followed monthly from September to April to investigate if there is an overall difference in the proportion of subjects who get sick between two treatment groups. A baseline of 60% disease rate for the common cold is used based on previous studies. Investigators would like to detect a treatment to placebo odds ratio of 0.5, which corresponds to a treatment group disease rate is 42.9%. Correspondingly, we have $\beta_1 = 0.405$, and $\beta_2 = -0.691$.

We would like to calculate the sample size for the study above with type I error $\alpha = 0.05$ and power $1 - \gamma = 0.8$ under a balanced design ($\bar{r} = 0.5$). We set the measurement times at $t_j = j - 1$ ($j = 1, \dots, 7$). We assume an AR(1) within-subject correlation structure with $\rho = 0.5$. The observation probability is $(\delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6, \delta_7) = (1, 0.95, 0.9, 0.85, 0.8, 0.75, 0.7)$. The sample sizes required under the IM, MM, and MIX (assuming a balanced mixture of IM and MM) patterns are 102, 108, and 105, respectively. On the other hand, when the within-subject correlation structure is CS, the other design parameters being the same, the required sample sizes under the IM, MM, and MIX patterns are 162, 172, and 167, respectively.

6 Discussion

In this study, we present a closed-form sample size formula for the test of TAD between two intervention groups in clinical trials with repeated binary outcomes. This sample size formula is flexible enough to account for arbitrary missing patterns and within-subject

correlation structures. In particular, we demonstrate that the proposed sample size can accommodate a mixture of missing patterns, which is frequently encountered by practitioners in real trials. To our knowledge, this is the first study that considers the mixture of missing patterns in sample size calculation. Our simulation shows that the nominal power and type I error are preserved over a wide range of sample sizes.

One limitation of the proposed sample size formula is that it is derived under the MCAR assumption. Under the MAR (missing at random) or MNAR (missing not at random) assumptions, however, an additional model is usually required to account for the missing mechanism. Because missing mechanisms are often different across clinical trials, it is almost impossible to provide a general sample size formula, let alone one with a closed form.

In this paper we have derived the sample size based on the assumption of a constant treatment effect (β_2). In reality, it is likely the treatment effect varies over time. Such a scenario can be accommodated by an extension of Equation (1), where **β_2 is replaced by β_{2j} . Then the vector of $\beta_2 = (\beta_{21}, \dots, \beta_{2m})'$** represents the variation of treatment effect over time $t_j (j = 1, \dots, m)$. For sample size calculation with respect to the test of TAD, it can be shown that this change in the assumption of treatment effect only impacts the denominator of Equation (8), where **β_{20} is replaced by an averaged value:**

$$\bar{\beta}_{20} = \left(\sum_{j=1}^m \delta_j / \beta_{2j0} \right) / \sum_{j=1}^m \delta_j.$$
 Here β_{20} is the true treatment effect at t_j . The numerator of Equation (8) remains unchanged. Because we denote the treatment effect by $\beta_2 = (\beta_{21}, \dots, \beta_{2m})'$, this approach is flexible enough to accommodate arbitrary trends in treatment effect.

The sample size presented in this paper is a natural extension from Zhang & Ahn (2012), which investigated sample size calculation for the test of TAD in clinical trials with repeatedly measured continuous outcomes. Detailed discussion about the impact of correlation and missing data on sample size can be found in Zhang & Ahn (2010). Recently there has been some new development in the field of sample size determination for clinical trials with longitudinal measurements. For example, Lu et al. (2009) proposed sample size determination for constrained longitudinal data analysis, where the baseline mean responses are constrained to be the same across treatment groups due to randomization. Lu (2012) investigated sample size calculations with multiplicity adjustment for longitudinal clinical trials with missing data. Both methods were developed in the context of continuous outcomes. Investigating the extension of such methods to clinical trials with binary outcomes will be one of our future research topics.

Acknowledgment

The work was supported in part by NIH grant 1UL1TR001105, AHRQ grant R24HS22418, and CPRIT grants RP110562-C1 and RP120670-C1.

7 Appendix

Appendix A. Proof of Theorem 1

We separate $A_n(\hat{\mathbf{b}})$ and $\Sigma_n(\hat{\mathbf{b}})$ into two parts (control and treatment),

$$\begin{aligned}
 A_n(\hat{\mathbf{b}}) &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \Delta_{ij} p_{ij}(\hat{\mathbf{b}}) q_{ij}(\hat{\mathbf{b}}) \begin{pmatrix} 1 & r_i - \bar{r} \\ r_i - \bar{r} & (r_i - \bar{r})^2 \end{pmatrix} \\
 &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \Delta_{ij} p_{ij}(\mathbf{b}) q_{ij}(\mathbf{b}) \begin{pmatrix} 1 & r_i - \bar{r} \\ r_i - \bar{r} & (r_i - \bar{r})^2 \end{pmatrix} + o_p(1) \\
 &= \frac{n_1}{n} \left\{ \frac{1}{n_1} \sum_{i=1}^{n_1} \sum_{j=1}^m \Delta_{ij} p_1 q_1 \begin{pmatrix} 1 & -\bar{r} \\ -\bar{r} & (\bar{r})^2 \end{pmatrix} \right\} + \frac{n_2}{n} \left\{ \frac{1}{n_2} \sum_{i=n_1+1}^n \sum_{j=1}^m \Delta_{ij} p_2 q_2 \begin{pmatrix} 1 & 1-\bar{r} \\ 1-\bar{r} & (1-\bar{r})^2 \end{pmatrix} \right\} + o_p(1)
 \end{aligned}
 \tag{10}$$

and

$$\begin{aligned}
 \Sigma_n(\hat{\mathbf{b}}) &= \frac{1}{n} \sum_{i=1}^n \left\{ \sum_{j=1}^m \Delta_{ij} \hat{\epsilon}_{ij} Z_{ij} \right\}^{\otimes 2} \\
 &= \frac{1}{n} \sum_{i=1}^n \left\{ \sum_{j=1}^m \Delta_{ij} (y_{ij} - p_{ij}) \begin{pmatrix} 1 \\ r_i - \bar{r} \end{pmatrix} \right\}^{\otimes 2} + o_p(1) \\
 &= \frac{n_1}{n} \left\{ \frac{1}{n_1} \sum_{i=1}^{n_1} \left\{ \sum_{j=1}^m \Delta_{ij} (y_{ij} - p_1) \begin{pmatrix} 1 \\ -\bar{r} \end{pmatrix} \right\}^{\otimes 2} \right\} + \frac{n_2}{n} \left\{ \frac{1}{n_2} \sum_{i=n_2+1}^n \left\{ \sum_{j=1}^m \Delta_{ij} (y_{ij} - p_2) \begin{pmatrix} 1 \\ 1-\bar{r} \end{pmatrix} \right\}^{\otimes 2} \right\} + o_p(1),
 \end{aligned}
 \tag{11}$$

Applying the central limit theorem to Equations (10) and (11), they converge to

$$\mathbf{A} = (1-\bar{r}) p_1 q_1 \sum_{j=1}^m \delta_j \begin{pmatrix} 1 & -\bar{r} \\ -\bar{r} & \bar{r}^2 \end{pmatrix} + \bar{r} p_2 q_2 \sum_{j=1}^m \delta_j \begin{pmatrix} 1 & 1-\bar{r} \\ 1-\bar{r} & (1-\bar{r})^2 \end{pmatrix} \tag{12}$$

and

$$\Sigma = \begin{pmatrix} 1 - \bar{r} \\ -\bar{r} \end{pmatrix} p_1 q_1 \sum_{j=1}^m \sum_{j'=1}^m \delta_{jj'} \rho_{jj'} \begin{pmatrix} 1 & -\bar{r} \\ -\bar{r} & \bar{r}^2 \end{pmatrix} + \bar{r} p_2 q_2 \sum_{j=1}^m \sum_{j'=1}^m \delta_{jj'} \rho_{jj'} \begin{pmatrix} 1 & 1 - \bar{r} \\ 1 - \bar{r} & (1 - \bar{r})^2 \end{pmatrix} \quad (13)$$

respectively, when $n \rightarrow \infty$.

After some algebra, it can be shown that the (2, 2)*th* element of $V = A^{-1} \Sigma A^{-1}$ is

$$\sigma_2^2 = \frac{\tau \sum_{j=1}^m \sum_{j'=1}^m \delta_{jj'} \rho_{jj'}}{\left(\sum_{j=1}^m \delta_j \right)^2 \sigma_r^2 p_1 q_1 p_2 q_2},$$

where $\tau = (1 - \bar{r}) p_1 q_1 + \bar{r} p_2 q_2$ and $\sigma_r^2 = \bar{r} (1 - \bar{r})$.

References

- Diggle, P., Heagerty, P., Liang, K., Zeger, S. Analysis of longitudinal data. 2nd ed.. Oxford University Press; 2002.
- Emrich LJ, Piedmonte M. Method for generating high-dimensional multivariate binary variates. *American Statistician*. 1991; 5:302–304.
- Hintze, J. PASS 13. NCSS, LLC; Kaysville, Utah, USA: 2013. www.ncss.com
- Jung S, Ahn C. Sample size for a two-group comparison of repeated binary measurements using GEE. *Statistics in Medicine*. 2005; 24(17):2583–2596. [PubMed: 16118812]
- Liang K, Zeger S. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; 73(1):13–22.
- Lipsitz S, Fitzmaurice G. Sample size for repeated-measurement studies with binary responses. *Statistics in Medicine*. 1994; 13:1233–1239. [PubMed: 7973204]
- Liu G, Liang K-Y. Sample size estimation for studies with correlated observations. *Biometrics*. 1997; 53:937–947. [PubMed: 9290224]
- Lu K. Sample size calculations with multiplicity adjustment for longitudinal clinical trials with missing data. *Statistics in Medicine*. 2012; 31(1):19–28. [PubMed: 22162151]
- Lu K, Mehrotra D, Liu G. Sample size determination for constrained longitudinal data analysis. *Statistics in Medicine*. 2009; 28(4):679–699. [PubMed: 19051207]
- Lui K-J. Sample sizes for repeated measurements in dichotomous data. *Statistics in Medicine*. 1991; 10(3):463–472. [PubMed: 2028129]
- Overall J, Doyle S. Estimating sample sizes for repeated measurement designs. *Controlled Clinical Trials*. 1994; 15(2):100–123. [PubMed: 8205802]
- Patel HI, Rowe E. Sample size for comparing linear growth curves. *Journal of Biopharmaceutical Statistics*. 1999; 9(2):339–350. [PubMed: 10379697]
- Zeger S, Liang K. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*. 1986; 42(1):121–130. [PubMed: 3719049]
- Zhang S, Ahn C. Effects of correlation and missing data on sample size estimation in longitudinal clinical trials. *Pharmaceutical Statistics*. 2010; 9(1):2–9. [PubMed: 19053162]

Zhang S, Ahn C. Sample size calculation for time-averaged differences in the presence of missing data. *Contemporary Clinical Trials*. 2012; 33(3):550–556. [PubMed: 22553832]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Sample Size (Empirical Power, Empirical Type I error) under fixed measurement time with $p_1 = 0.5$

δ	CS		AR(1)	
	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.3$	$\rho = 0.5$
IM				
δ_1	216(0.794,0.053)	303(0.792,0.045)	143(0.794,0.049)	203(0.799,0.052)
δ_2	229(0.802,0.050)	315(0.792,0.044)	156(0.795,0.052)	216(0.797,0.054)
δ_3	225(0.794,0.051)	311(0.795,0.042)	153(0.798,0.056)	213(0.802,0.052)
δ_4	232(0.807,0.055)	319(0.800,0.053)	159(0.793,0.056)	218(0.797,0.050)
MM				
δ_2	237(0.800,0.052)	330(0.796,0.056)	161(0.786,0.048)	226(0.805,0.051)
δ_3	229(0.803,0.049)	318(0.796,0.054)	156(0.797,0.050)	219(0.803,0.046)
δ_4	246(0.787,0.053)	342(0.795,0.051)	167(0.795,0.049)	234(0.802,0.054)
MIX				
δ_2	233(0.799,0.045)	322(0.798,0.045)	159(0.794,0.042)	221(0.796,0.043)
δ_3	227(0.797,0.053)	315(0.801,0.049)	154(0.792,0.042)	216(0.787,0.044)
δ_4	239(0.788,0.044)	330(0.795,0.042)	163(0.793,0.042)	226(0.796,0.041)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Sample Size (Empirical Power, Empirical Type I error) under fixed measurement time with $p_1 = 0.2$

δ	CS		AR(1)	
	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.3$	$\rho = 0.5$
IM				
δ_1	291(0.792,0.050)	407(0.801,0.051)	193(0.791,0.055)	273(0.799,0.056)
δ_2	307(0.797,0.052)	423(0.797,0.050)	210(0.799,0.049)	290(0.806,0.049)
δ_3	303(0.799,0.056)	419(0.803,0.054)	206(0.800,0.049)	287(0.796,0.056)
δ_4	313(0.803,0.049)	429(0.802,0.057)	214(0.794,0.058)	293(0.801,0.052)
MM				
δ_2	319(0.789,0.054)	443(0.801,0.052)	217(0.806,0.052)	304(0.797,0.052)
δ_3	308(0.793,0.048)	428(0.797,0.057)	210(0.788,0.052)	294(0.786,0.054)
δ_4	331(0.790,0.056)	460(0.788,0.053)	225(0.793,0.053)	315(0.795,0.060)
MIX				
δ_2	313(0.782,0.046)	433(0.800,0.057)	213(0.788,0.050)	297(0.795,0.049)
δ_3	305(0.801,0.048)	423(0.796,0.046)	208(0.794,0.040)	290(0.791,0.047)
δ_4	322(0.795,0.044)	444(0.799,0.051)	219(0.795,0.044)	304(0.809,0.046)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript